

Microwave Spectroscopy Based Classification of Rat Hepatic Tissues: On the Significance of Dataset

T. YILMAZ

Abstract—With the advancements in machine learning (ML) algorithms, microwave dielectric spectroscopy emerged as a potential new technology for biological tissue and material categorization. Recent studies reported the successful utilization of dielectric properties and Cole-Cole parameters. However, the role of the dataset was not investigated. Particularly, both dielectric properties and Cole-Cole parameters are derived from the S parameter response. This work investigates the possibility of using S parameters as a dataset to categorize the rat hepatic tissues into cirrhosis, malignant, and healthy categories. Using S parameters can potentially remove the need to derive the dielectric properties and enable the utilization of microwave structures such as narrow or wideband antennas or resonators. To this end, *in vivo* dielectric properties and S parameters collected from hepatic tissues were classified using logistic regression (LR) and adaptive boosting (AdaBoost) algorithms. Cole-Cole parameters and a reproduced dielectric property data set were also investigated. Data preprocessing is performed by using standardization a principal component analysis (PCA). Using the AdaBoost algorithm over 93% and 88% accuracy is obtained for dielectric properties and S parameters, respectively. These results indicate that the classification can be performed with a 5% accuracy decrease indicating that S parameters can be an alternative dataset for tissue classification.

Index Terms—Cole-Cole parameters, dielectric properties, *in vivo* measurements, machine learning, rat hepatic tissues.

I. INTRODUCTION

DIELECTRIC PROPERTY discrepancy between healthy and diseased biological tissues enabled many different applications of electromagnetics in medicine including but not limited to microwave hyperthermia, microwave breast cancer imaging, and microwave blood glucose detection [1-3]. Microwave dielectric properties can be characterized using different methods. The characterization method is determined based on different features including the frequency of operation, nature of sample, and temperature range. One such technique that has been used for broadband dielectric property measurements of high permittivity and high loss materials is the open-ended coaxial probe technique. Open-ended coaxial probes are commercially available for laboratory use; therefore,


the technique has been widely utilized for dielectric property characterization of different materials including liquids, gel-like materials, along with biological tissues [4]. Being a broadband method, it is known that the technique suffers from large error rates. The source of error can be user, sample, technique based or a combination of those.

It is recommended that during the laboratory use, the user should follow a calibration scheme before measurement and the calibration should be repeated after completion of a number of measurements, measurement set-up should be fixed and sample should be brought to the tip of the probe and tight contact between the probe and the measurement sample should be ensured. Even when the listed conditions are met during a measurement, the commercial probes reports $\pm 5\%$ error [5].

In the reported literature, despite the reported high measurement error, different potential applications of the technique have been proposed. An example of a practical application is the utilization of the probe for kidney stone classification [6]. Another practical realization is using the probe for biopsy or surgical margin determination or a biopsy probe [7,8]. In a practical setting, the error rate is expected to increase up to 30% [4]. To mitigate the error, several approaches can be applied including hardware and mathematical updates. However, previously reported studies proved that, without costly updates on the method, the accuracy of tissue classification can be increased by adopting machine learning (ML) based classification algorithms [6-8]. Such algorithms are not necessarily concerned with determining the dielectric properties of the samples; however, they determine the type of the sample under test. Based on the reported classification results in previous studies, the accuracy can be improved approximately 10% to 30% with this approach.

The reported accuracy advancements were obtained by application of classification algorithms to two types of datasets the first and mostly used one is the dielectric property data and the second one is the parameters of mathematical models [6-8]. The mathematical models are widely used to represent the dispersive dielectric property behavior with respect to frequency. The goal is to use fewer parameters that can later be generalized to obtain dielectric properties at the desired frequencies. Mathematical models have been widely used in the

TUBA YILMAZ, is with Department of Electronics and Communication Engineering, Istanbul Technical University, Istanbul, Turkey, (e-mail: tuba.yilmaz@itu.edu.tr).

 <https://orcid.org/0000-0003-3052-2945>

Manuscript received July 28, 2020; accepted July 28, 2020.

DOI: [10.17694/bajece.419538](https://doi.org/10.17694/bajece.419538)

literature to report the measured dielectric properties of biological tissues as well as several different liquids [9]. One of the commonly used mathematical model is the single pole Cole-Cole equation.

Both dielectric properties and Cole-Cole parameters are calculated parameters. The dielectric properties are calculated from the S parameters which are directly measurable quantities. The Cole-Cole parameters are fitted to calculated dielectric properties. Although the performance of classification algorithms was evaluated for both dielectric properties and Cole-Cole parameters in the literature, classification with S parameters have not been explored. This work presents a comparison of the multi-class classification results based on four datasets, namely dielectric properties, Cole-Cole parameters, S parameters, and reproduced dielectric properties from Cole-Cole parameters. All data sets are obtained from same rat hepatic tissue samples and in some cases, they are reproduced from one another. Classification with these datasets investigated to explore the role of the data type on classification accuracy. More significantly this work investigates S parameter-based classification that can potentially eliminate the need for dielectric property retrieval and Cole-Cole parameter calculation.

The remainder of this paper is organized as follows: In section II the measurement set-up, *in vivo* data collection, data curation and the ML algorithms are detailed, the raw data for each group, data obtained through modelling and classification results for each data type are given in section III, conclusions are drawn in section IV.

II. METHODOLOGY

A detailed description of the dielectric property measurement, as well as the description of samples, were given in [7-10]. Therefore, we briefly explain the sample preparation and dielectric property measurements in this section.

A. Measurement Set-up and Calibration

The measurement set-up included an open-ended coaxial probe connected to a Fieldfox N9923A Network Analyzer (NA) with an RF cable. Type of the probe was a slim-form with 2.2 mm aperture diameter. Agilent 85070 software was utilized for dielectric property data collection. The calibration was performed by following standard open-short-deionized water calibration procedure. Deionized water temperature was measured before the calibration and entered to the software. The probe tip was air dried after completion of the calibration.

B. In vivo Data Collection

Female Wistar Albino rats were obtained from the Istanbul University, Institute of Experimental Medicine at 120 days old. The number of experiment animals were 30. The animals were divided into two groups namely control and experiment. The experiment group received 50 (mg/kg) Diethylnitrosamine (Sigma Chemical Company, St. Louis, MO, USA)- solution once a week via intraperitoneal injection. The control group received an intraperitoneal injection of 0.1M saline solution. After 10 weeks the animals were left to rest for 6 weeks.

Throughout the experiments the animals were kept in 12-hour light/dark cycle and had ad libitum access to tap water and

standard pellet food. During the chemical induction period of 10 weeks, 8 animals died from the experiment group. After the rest period is completed measurements were taken and the animals were immediately sacrificed upon completion of the measurements. Before measurements the animals were anesthetized via intraperitoneal injection of 80 (mg/kg) ketamine + 10 (mg/kg) xylazine mixture. The experiments were in accordance with the Istanbul University, Animal Experiments Local Research Ethics Committee. To collect the measurements, anesthetized animal's liver is accessed via intraperitoneal incision. Two sets of measurements were taken; that is, wet and dry. The wet measurements were taken immediately after incision. Dry measurements were taken after the wet measurement area is wiped with a 0.1M saline solution. This procedure is performed to clean the area from accumulated blood. The procedure is also widely practiced during surgeries. The number of collected measurements varied based on the sample availability.

C. Cole-Cole Fitting

Cole-Cole equation is widely used in the literature to represent the dispersive dielectric property behavior of many different materials including but not limited to biological tissues, biological fluids, and polar liquids such as alcohols. The variables in Cole-Cole equation are $\Delta\epsilon$ dielectric constant difference between lower and higher frequencies ($\Delta\epsilon = \epsilon_s - \epsilon_\infty$), τ is the relaxation time, α relaxation constant, and σ_i ionic conductivity. These variables are named Cole-Cole parameters. Cole-Cole parameters are fitted to the measured dielectric properties with various curve fitting techniques. In this work, the parameters were fitted with Particle Swarm Optimization (PSO) algorithm, the details of the technique were described in [6,10]; therefore, it is briefly given in this article. The Cole-Cole equation is given below,

$$\epsilon' - i\epsilon'' = \epsilon_\infty + \frac{\Delta\epsilon}{1 + (i\omega\tau)^{(1-\alpha)}} + \frac{\sigma_i}{i\omega\epsilon_0} \quad (1)$$

where ϵ' is dielectric constant, ϵ'' is dielectric loss factor, ω is angular frequency ($\omega = 2\pi f$). The PSO algorithm searches a pre-determined solution space to reach an optimal solution. The pre-determined solution domain is constructed by defining intervals to the listed five Cole-Cole parameters. The intervals are given in Table I. PSO is a stochastic optimization technique developed based on the swarm behavior. In a swarm each member's next location is determined through a combination of the experience of the individual and the swarm. These individuals are named particles and in this work their locations, namely the Cole-Cole parameters, are kept as a vector. 20 particles were used in this work, the particles evaluate their locations on each iteration and determine the next location based on the best value obtained by all particles and the best value obtained by each particle. The goodness of a location is determined by reproducing dielectric properties between 0.5 to 6 GHz with 0.5 GHz resolution and finding the Euclidean

distance between the reproduced and measured dielectric properties. The Euclidean distance is given with equation (2),

$$error = \frac{1}{N} \sum_{i=1}^N \left[\left(\frac{\varepsilon'_{\omega_i} - \hat{\varepsilon}'_{\omega_i}}{\text{median}(\varepsilon'_{\omega_i})} \right)^2 - \left(\frac{\varepsilon''_{\omega_i} - \hat{\varepsilon}''_{\omega_i}}{\text{median}(\varepsilon''_{\omega_i})} \right)^2 \right] \quad (2)$$

N is the number of frequency points used during the measurements (N=12). ε'_{ω_i} and ε''_{ω_i} represents the measured dielectric constant and loss factor, respectively. $\hat{\varepsilon}'_{\omega_i}$ and $\hat{\varepsilon}''_{\omega_i}$ represents the calculated dielectric constant and loss factor, respectively. Maximum iteration number is set to 50. If a solution reached a Euclidean distance smaller than the threshold of 0.001 otherwise the algorithm keeps trying to find the best location until the iteration limit is reached.

TABLE I
COLE-COLE PARAMETER SOLUTION INTERVALS

Cole-Cole Parameters	ε_{∞}	$\Delta\varepsilon$	τ (ps)	α	σ_i (S/m)
Min	1	1	1	0.01	0.01
Max	80	80	20	1	3

After fitting the Cole-Cole parameters a new dataset was reproduced from the fitted Cole-Cole values. This step was taken in order to reduce dielectric property discrepancy between the frequency points. Though smoothing the dielectric properties with respect to frequency, the measurement errors can be potentially mitigated. The Cole-Cole fitting results as well as the reproduced data is given in Section III.

D. Data Pre-processing

Four sets of data obtained from dielectric properties, S parameter measurements, Cole-Cole parameters and reproduced dielectric properties were used in three different forms to train and test the ML algorithms. The raw data is used without pre-processing.

The second set was the standardized datasets. Standardization is known to re-distribute the data to resemble a Gaussian distribution. Some algorithms are known to perform well on standardized data such as linear learners. In fact, standardization is a requirement for many ML algorithms. The standardization is performed on each feature by subtracting the mean (u) from each data point (x) and dividing it by the standard deviation (s) ($z=(x-u)/s$).

Third kind of data sets were obtained by applying principal component analysis (PCA) to the standardized data. PCA is used for dimensionality reduction purposes and to transform the data in order to maximize the discrepancy between the classes [11]. In this work, we used the PCA to exploit the difference between the classes since the dimensionality reduction was not a concern. PCA works by calculating new set of features, called eigenvalues, computed from the covariance matrix. This matrix describes the correlations between features. Eigenvalues does not represent a physically meaningful quantity; however, geometrically it represents the independent dimensions with

maximum variances. Both the PCA and standardization were applied separately to the train and test groups to avoid peeking.

E. Machine Learning Algorithms

Performances of two ML algorithms namely logistic regression (LR) and adaptive boosting (AdaBoost) were investigated. These algorithms were selected based on the previously reported research since the dielectric properties were successfully classified using the linear and ensemble methods. Both LR and AdaBoost algorithm designed for binary classification problems. One-vs-Rest (OvR) scheme is used to adopt the algorithms for multiclass classification. OvR is the most commonly used scheme for multiclass classification. It works by separating the original data to number of classes and applies the classifier to each set. One drawback is the computational cost specifically if the number of classes are large.

LR, a linear algorithm, performs the classification based on the calculated probabilities [12]. LR uses sigmoid function to estimate the probability of a sample belonging to one of two classes. Sigmoid function is given in equation (3).

$$p = \frac{1}{1 + e^x} \quad (3)$$

Sigmoid function is an S shaped curve limited between 0 and 1. Using maximum likelihood (ML) during training the LR algorithm adjusts the curve to obtain the optimum fitting. This work uses the L2 regularized cost function to solve the classification problem. One advantage of the LR algorithm is the ability to decrease the weights of the irrelevant features. That is, the algorithms tune itself to increase the weight of significant features. In earlier studies it was proven that the linear methods tend to perform well on dielectric property data classification.

AdaBoost algorithm is an ensemble method and uses weak classifiers to obtain a strong classifier [13]. This work used decision tree classifiers as weak classifiers. Typically, AdaBoost starts by training a weak classifier with the given training set by giving a weight of 1 to each data on the training set. Next, AdaBoost increases the weight of the misclassified training data and trains another classifier with the weighted training set. This process iteratively continues until the maximum iteration number is reached; the iteration number is set to 100 for this work. The trained classifiers receive a score and base on the score weighted combination of the classifiers is linearly combined to form a strong classifier.

Pre-processing and classification of the data are performed in Python language using scikit-learn libraries [14].

III. RESULTS

A. Dielectric properties

Total measurements collected from healthy, malignant and cirrhosis tissues are 395, 285, and 95, respectively. To form balanced classes the number of data points are deduced to 95 for healthy and malignant tissue classes. This is performed by randomly selecting data points from the two data points.

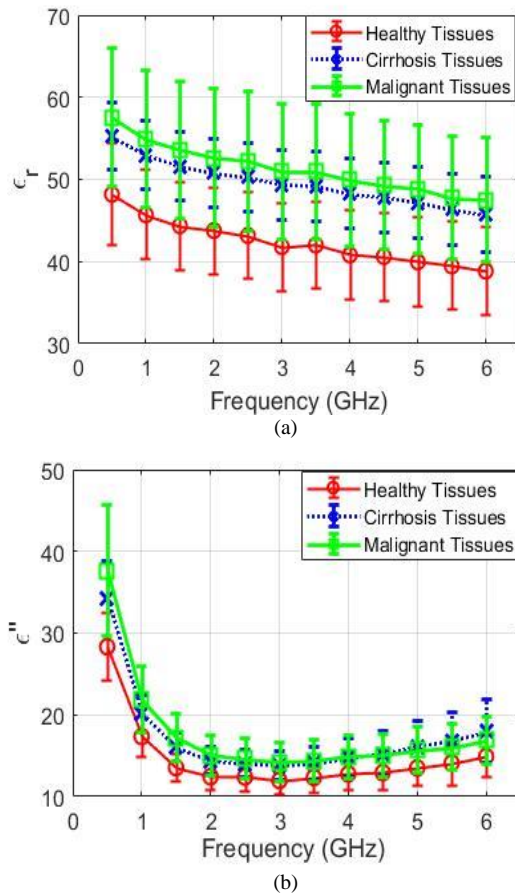


Fig.1. Mean and standard deviation of *in vivo* dielectric properties of healthy, cirrhotic, and malignant rat hepatic tissues; (a) Relative dielectric constant comparison, (b) Dielectric loss comparison.

The measurements are performed between 0.5 GHz to 6 GHz with 0.5 GHz resolution. It should be noted that in this work the whole data set is used and the measurements are not categorized as dry and wet. Mean as well as the standard deviation of dielectric constant and dielectric loss is given in Figure 1(a) and Figure 1(b), respectively, for each of the balanced tissue classes. Mean dielectric constant and loss discrepancy between healthy and malignant, healthy and cirrhosis, cirrhosis and malignant tissues are less than 21.2% and 19.9%, 17% and 17.3%, 3.6% and 4.6% for the whole frequency band, respectively.

B. S-parameter Measurements

The S parameters associated with the dielectric properties were also recorded during measurements. As noted, before, dielectric properties are essentially derived from the S parameter measurements. Figure 2(a) and Figure 2(b) shows the measured real and imaginary parts of S_{11} parameters corresponding to dielectric properties given with Figure 1.

Percent difference for mean real and imaginary parts of S_{11} response between healthy and malignant, healthy and cirrhosis, cirrhosis and malignant tissues are less than 20.5% and 11.1%, 18% and 9.1%, 3.4% and 1.8% for the whole frequency band, respectively. The percent discrepancies are slightly lower than dielectric property discrepancies.

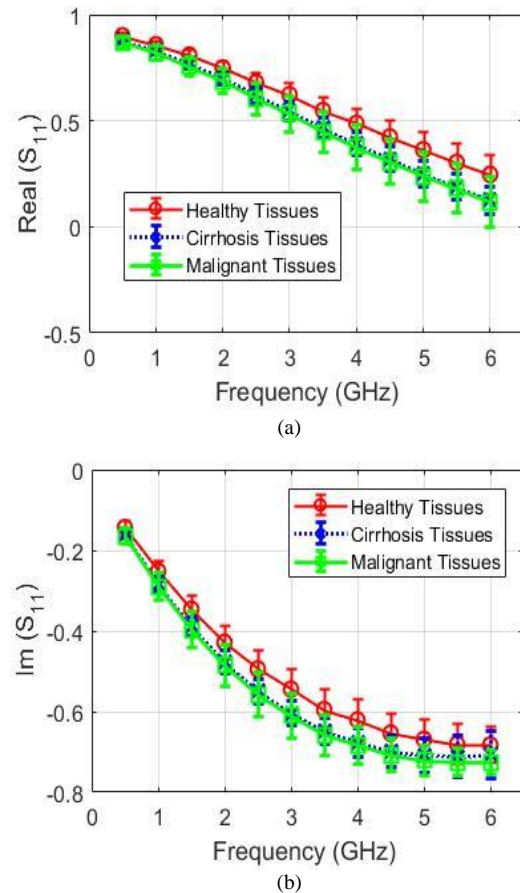
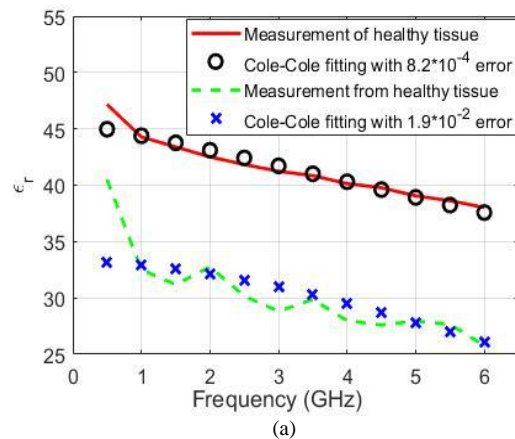


Fig.2. Mean and standard deviation of *in vivo* measured S_{11} response of the probe when the open end is terminated with healthy, cirrhotic, and malignant rat hepatic tissues; (a) Real part of S_{11} response comparison, (b) Imaginary part of S_{11} response comparison.

C. Cole-Cole parameters

Cole-Cole parameters were fitted to each 95 measurements collected from healthy, cirrhosis, and malignant rat tissue samples. Comparisons of sample measurements collected from healthy rat hepatic tissues with the fitted Cole-Cole parameters are shown in Figure 3.



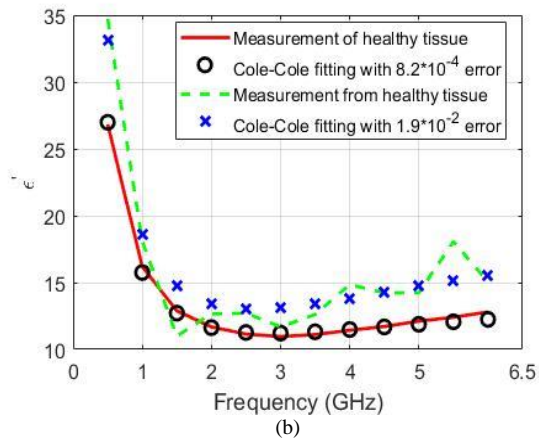


Fig.3. Comparison of measured dielectric properties collected from healthy rat hepatic tissues and the Cole-Cole fittings with minimum and maximum error; (a) Two dielectric constants and the corresponding Cole-Cole fittings, (b) Two dielectric loss and the corresponding Cole-Cole fittings

TABLE II
SAMPLES OF THE FITTED COLE-COLE PARAMETERS TO THE MEASURED DIELECTRIC PROPERTIES OF EACH GROUP WITH MAXIMUM AND MINIMUM ERROR RATES

Cole-Cole Parameters	ϵ_∞	$\Delta\epsilon$	τ (ps)	α	σ_i (S/m)	error
Healthy	15.37	30.15	12.51	0.15	0.7	8.2e-04
	1.63	31.56	14.05	0.01	0.88	1.9e-02
Cirrhosis	14.7	51.01	10.16	0.04	1.31	4.8e-04
	7.72	42.07	14.67	0.01	0.77	4.3e-03
Malignant	12.15	57.29	7.67	0.09	1.5	5.5e-04
	25.4	41.74	13.89	0.32	1.09	8.1e-03

Figure 3 (a) and Figure 3 (b) gives the dielectric constant and loss values, respectively, obtained from measurements and Cole-Cole fittings with maximum and minimum errors. As it can be seen from Figure 3, Cole-Cole fitting errors depends on the quality of the measurement.

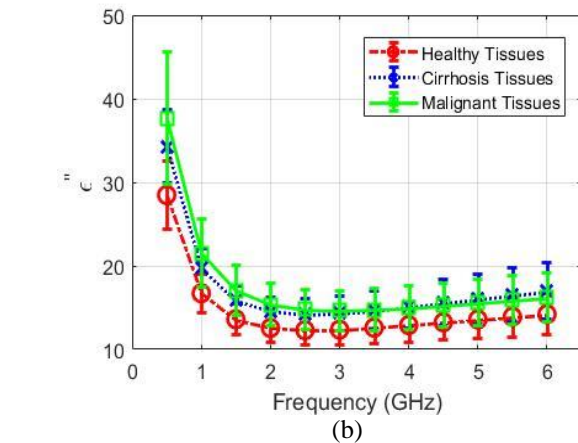
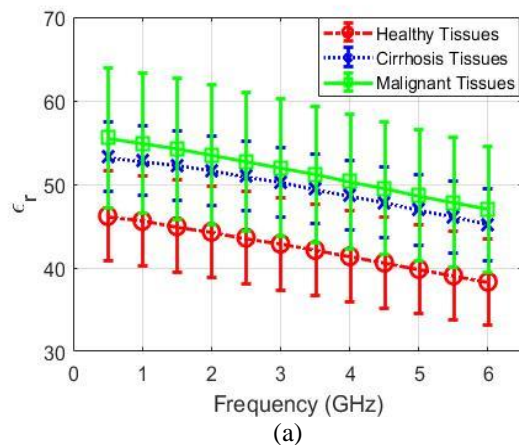


Fig.4. Mean and standard deviation of dielectric properties of healthy, cirrhotic, and malignant rat hepatic tissues obtained from the Cole-Cole fittings; (a) Relative dielectric constant comparison, (b) Dielectric loss comparison.

Fitted Cole-Cole parameters with maximum and minimum error for each tissue class is given in Table II. Fitted Cole-Cole parameters were used to produce dielectric property dataset. The goal was to reproduce the dielectric properties to minimize the measurement error. Mean and standard deviation of the reproduced dielectric property dataset for relative dielectric constant and dielectric loss are given in Figure 4(a) and Figure 4(b), respectively. Mean dielectric constant and loss discrepancy between healthy and malignant, healthy and cirrhosis, cirrhosis and malignant tissues are less than 21.3% and 19.9%, 17.0% and 17.5%, 3.7% and 4.3% for the whole frequency band, respectively.

D. Classification Performances

Leave-one-out (LOO) cross validation (CV) scheme was used for classification due to the limited number of the data (95 data from each class). Over 80% accuracy was obtained using LR algorithm along with raw dielectric properties. LR did not perform as well on raw S_{11} data set. LR seems to be sensitive to standardization since the accuracy loss reached up to 10 % for raw dielectric property data and it improved over 10% for S_{11} data when applied after the standardization. AdaBoost on the other hand is not sensitive to the standardization of the data the accuracies did not significantly changed after standardization. Both Cole-Cole parameters and the reproduced dielectric properties from Cole-Cole parameter datasets did not perform well with both algorithms. This indicates that despite the low error fittings part of the data is corrupted during the process. It should also be noted that the classification with Cole-Cole parameters can be achieved with high accuracy when the dielectric property data are collected under ideal conditions with low standard deviation [6].

AdaBoost performs with over 93% accuracy after the PCS is applied to the standardized dielectric property data. Similarly, AdaBoost is able to achieve 88% accuracy after the application of PCA to standardized S_{11} data. Having said that, hyperparameter optimization was not performed in this work. We can state that further potential improvement of the accuracy results is possible with hyperparameter optimization.

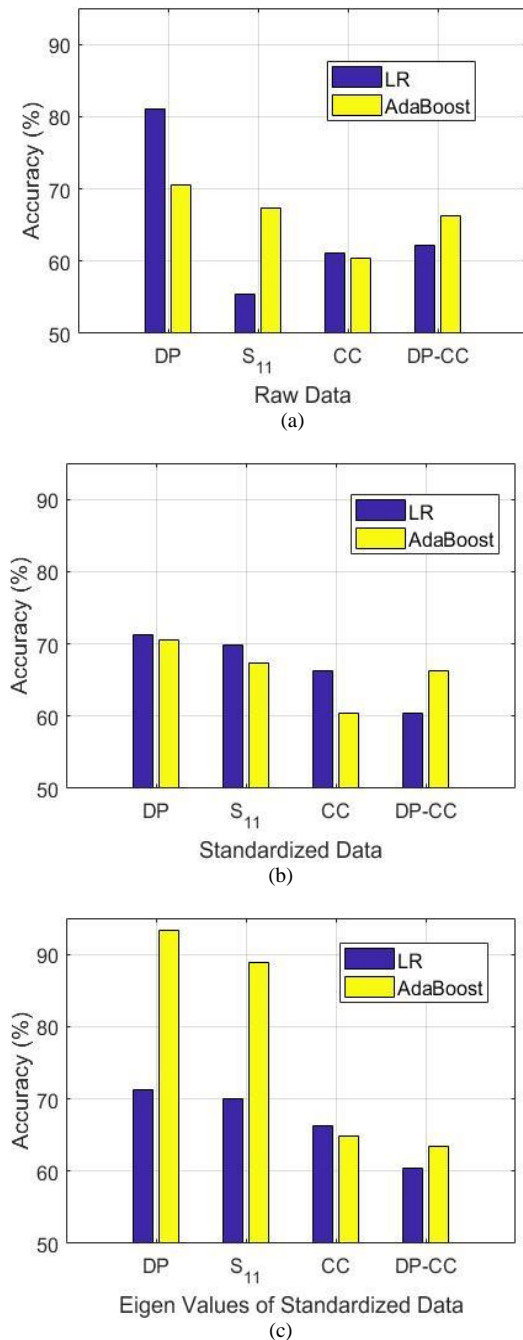


Figure 5. Accuracy results obtained from application of Logistic Regression (LR) and Adaptive Boosting (AdaBoost) algorithms to dielectric properties (DP), measured S_{11} parameters, Cole-Cole (CC) parameters fitted to the DP, and DP reproduced by using the fitted CC (DP-CC): (a) Accuracy results obtained from the raw data, (b) Accuracy results obtained from the standardized data, (c) Accuracy results obtained through the application of Principal Component Analysis to the standardized data.

IV. CONCLUSIONS

This work investigated the significance of dataset for microwave measurement-based classification of the rat hepatic malignancies. In the literature, different microwave spectroscopy datasets were utilized to classify different biological tissue types and materials. Dielectric properties, S parameters, fitted Cole-Cole parameters, and dielectric

properties reproduced from the Cole-Cole parameters were investigated to understand the classification accuracy performances. Two common classifiers LR and AdaBoost were used. Note that the classification based on S_{11} parameters have not been investigated in the literature. It was concluded that the classification accuracies can be increased by employing PCA and ensemble methods tend to reach over 88% accuracies for S_{11} and dielectric property datasets. By using S_{11} the accuracy decreases by 5% indicating that S_{11} parameter-based classification can produce reliable results.

REFERENCES

- [1] T. U. Gürbüz, B. Aslanyürek, A. Yapar, H. Şahintürk I. Akduman. "A Nonlinear Microwave Breast Cancer Imaging Approach Through Realistic Body-Breast Modeling." *IEEE Transactions on Antennas and Propagation*, vol. 62, 5, 2014, pp. 2596-2605.
- [2] M. Converse, E. J. Bond, S. C. Hagness, B. D. Van Veen. "Ultrawide-band microwave space-time beamforming for hyperthermia treatment of breast cancer: a computational feasibility study." *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, 8, 2004, pp. 1876-1889.
- [3] T. Yilmaz, R. Foster, Y. Hao. "Radio-Frequency and Microwave Wave Techniques for Non-Invasive Measurement of Blood Glucose Levels." *Diagnostics*, vol. 9.1, 2019, pp. 1-34.
- [4] D. Popovic, L. McCartney, C. Beasley, M. Lazebnik, M. Okoniewski, S. C. Hagness, J. H. Booske. "Precision open-ended coaxial probes for in vivo and ex vivo dielectric spectroscopy of biological tissues at microwave frequencies." *IEEE Transactions on Microwave Theory and Techniques*, vol. 53.5, 2005, pp. 1713-1722.
- [5] Keysight Technologies. Probe Characteristics and Specifications, Keysight N1501A, Dielectric Probe Kit 10 MHz to 50 GHz. Available online: <https://literature.cdn.keysight.com/litweb/pdf/5992-0264EN.pdf?id=2605692> (accessed on 25 July 2020).
- [6] B. Saçlı, C. Aydınalp, G. Cansız, S. Joof, T. Yilmaz, M. Çayören, B. Önal, I. Akduman. "Microwave dielectric property based classification of renal calculi: Application of a kNN algorithm." *Computers in biology and medicine*, vol. 112, 2019, pp. 103366.
- [7] T. Yilmaz. "Multiclass Classification of Hepatic Anomalies with Dielectric Properties: From Phantom Materials to Rat Hepatic Tissues." *Sensors*, vol. 20, 2020, pp. 530.
- [8] T. Yilmaz, M. A. Kılıç, M. Erdoğan, M. Çayören, D. Tunaoğlu, İ. Kurtoglu, Y. Yaslan et al. "Machine learning aided diagnosis of hepatic malignancies through in vivo dielectric measurements with microwaves." *Physics in medicine & biology*, vol. 61.13, 2016, pp. 5089.
- [9] S. Gabriel, R. W. Lau, C. Gabriel. "The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues." *Physics in medicine & biology*, vol. 41.11, 1996, pp. 2271.
- [10] T. Yilmaz, F. Ates Alkan. "In Vivo Dielectric Properties of Healthy and Benign Rat Mammary Tissues from 500 MHz to 18 GHz." *Sensor*, vol. 20, pp. 2214.
- [11] T. Jolliffe, J. Cadima. "Principal component analysis: a review and recent developments." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, pp. 2065, 2016.
- [12] L. Shen, E.C. Tan. "Dimension reduction-based penalized logistic regression for cancer classification using microarray data." *IEEE/ACM Transactions on computational biology and bioinformatics*, vol. 2.2, 2005, pp. 166-175.
- [13] T. Hastie, R. Saharon, J. Zhu, H. Zou. "Multi-class adaboost." *Statistics and its Interface*, vol. 2.3, 2009, pp. 349-360.
- [14] Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, *JMLR* 12, 2011, pp. 2825-2830.